ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Development and Valuation of a Preference-Weighted Measure in Age-Related Macular Degeneration From the Vision Impairment in Low Luminance Questionnaire—A MACUSTAR Report

Donna Rowen, PhD, Jill Carlton, PhD, Jan H. Terheyden, MD, Robert P. Finger, MD, PhD, Nyantara Wickramasekera, MSc, John Brazier, PhD, on behalf of the MACUSTAR Consortium

ABSTRACT

Objectives: This study generates VILL-UI (Vision Impairment in Low Luminance - Utility Index), a preference-weighted measure (PWM) derived from the VILL-33 measure for use in patients with age-related macular degeneration (AMD) and valued to generate United Kingdom and German preference weights.

Methods: A PWM consists of a classification system to describe health and utility values for every state described by the classification. The classification was derived using existing data collected as part of the MACUSTAR study, a low-interventional study on AMD, conducted at 20 clinical sites across Europe. Items were selected using psychometric and Rasch analyses, published criteria around PWM suitability, alongside instrument developer views and concept elicitation work that informed VILL-33 development. An online discrete choice experiment (DCE) with duration of the health state was conducted with the United Kingdom and German public. Responses were modeled to generate utility values for all possible health states.

Results: The classification system has 5 items across the 3 domains of VILL-33: reading and accessing information, mobility and safety, and emotional well-being. The DCE samples (United Kingdom: n = 1004, Germany: n = 1008) are broadly representative and demonstrate good understanding of the tasks. The final DCE analyses produce logically consistent and significant coefficients.

Conclusions: This study enables responses to VILL-33 to be directly used to inform economic evaluation in AMD. The elicitation of preferences from both United Kingdom and Germany enables greater application of VILL-UI for economic evaluation throughout Europe. VILL-UI fills a gap in AMD in which generic preference-weighted measures typically lack sensitivity.

Keywords: age-related macular degeneration, preference elicitation, preference-weighted measure, quality-adjusted life-year, Vision Impairment in Low Luminance-33.

VALUE HEALTH. 2024; ■(■):■–■

Highlights

- The Vision Impairment in Low Luminance-33 (VILL-33) patient-reported outcome measure is psychometrically valid and captures vision-related quality of life in age-related macular degeneration but cannot be used in its current form to inform economic evaluation because it does not have preference weights.
- This study generates VILL-UI, a preference-weighted measure derived from the VILL-33 measure for use in patients with age-related macular degeneration, with both United Kingdom and German preference weights.
- This study generates a new preference-weighted measure VILL-UI, which enables prospective and retrospective data sets with VILL-33 data to be directly used to inform economic evaluation.

Introduction

Resource allocation decisions are being increasingly informed by cost-effectiveness modeling in which outcomes are represented using quality-adjusted life-years (QALYs). QALYs are generated by multiplying a quality adjustment weight by life-years, enabling capture of improvements and deteriorations in both health-related quality of life (HRQoL) and life expectancy to assess the cost-effectiveness of different treatments. QALYs are usually generated using a preference-weighted measure (PWM) to generate the quality adjustment weight, such as the EQ-5D, the most commonly used generic PWM of HRQoL.[1] The weight must be based on preferences to generate utilities, reflecting how good or bad the level of HRQoL is perceived to be.[2] However, generic measures do not focus on aspects of HRQoL that may be important for patients with a given condition. Vision strongly interferes with the ability to perform activities of daily living and is valued as the most relevant sense by the general population.[3] For conditions affecting vision, the EQ-5D-3L has been found to have poor psychometric performance,[4,5] and this is likely to also apply to the EQ-5D-5L.[6] Age-related macular degeneration (AMD), the most common blinding condition in industrialized countries, particularly affects central vision and impairs vision in low-luminance and low-contrast situations during its early stages. Similarly to vision impairment overall, the EQ-5D-3L was previously shown to perform poorly in AMD. Results of AMD studies using EQ-5D-3L suggest it can be unable to detect differences in severity subgroups despite these been reflected in condition-specific measures; hence, there is a concern that if used it may not accurately capture improvements or deteriorations in HRQoL due to treatment or progression,[7–14] although 1 study found it performed favorably.[15]

Condition-specific measures focus on what is important for patients with a condition. The vision impairment in low luminance (VILL) patient-reported outcome measure (PROM) has been recently developed for use in patients with AMD.[16] The VILL was developed to be able to capture the impact of difficulties with vision in low-luminance and low-contrast situations on HRQoL, which is required for use as an endpoint in early and intermediate AMD trials and has been found to be psychometrically sound in terms of repeatability, internal consistency, content, construct, and criterion-related validity.[17,18]

The VILL does not currently have a preference-weighted scoring, meaning that it cannot directly generate QALYs. One option to estimate utility values using VILL data is to map the measure to a PWM such as the EQ-5D-5L, to enable EQ-5D-5L utilities to be predicted using VILL items.[6,12] However, if the EQ-5D-5L and the VILL do not capture the same dimensions of HRQoL, the mapping will not be accurate, and the HRQoL aspects from the VILL that are important for patients are unlikely to be accurately represented in the mapped utility values.[19]

An alternative option is to generate an AMD-specific PWM from the VILL. A PWM consists of (1) a classification system to describe HRQoL and (2) utility values for every health state described by the classification system. A classification system comprises factors or dimensions, with 1 or more items within each dimension that best reflect the aspect of HRQoL captured by that dimension. A parsimonious number of items is required for the classification system to be amenable to valuation.

This aim of this study is to develop the VILL-UI (utility index), an AMD-specific PWM from the VILL-33, and generate preference weights for the United Kingdom and Germany using preferences from representative samples of the respective general populations. This will enable AMD-specific utilities to be generated from VILL data for use in economic analyses of interventions in AMD.

Methods

The development of an AMD-specific PWM from the VILL-33 comprises 2 stages: (1) selection of a subset of items from the VILL-33 to generate a health state classification system that can be used to describe the HRQoL of patients with AMD and (2) preference weights for the United Kingdom and Germany that enable health-state utility values to be generated for every health state described by the classification system.

Derivation of the VILL-UI Health-State Classification System From VILL for Valuation

VILL

The VILL PROM was developed using existing questionnaire items, in-depth interviews, focus group discussions, and cognitive debriefs with AMD patients.[16] Psychometric analyses (including investigation of item fit, internal consistency, person-item targeting, ordering of thresholds, differential item functioning, and dimensionality based on polytomous Rasch models) undertaken on the original version of the VILL, VILL-37, identified 4 of the 37 items as candidates for exclusion to generate the validated VILL-33.[17] The VILL-33 has 33 items across 3 subscales or factors: 17 items for "reading and accessing information," 12 items for "mobility and safety," and 4 items for "emotional well-being." Items have 4 response options reflecting difficulty (items 1-24, response options: can't do because of eyesight; a lot; a little; none) or frequency (items 25-33, response options: always; often; sometimes; never) with an additional response option stating that the item is not applicable (Didn't do this for other reasons/Does not apply to me).

VILL data

Data were collected as part of the MACUSTAR study, a low-interventional study on AMD, including patients with early, intermediate, and late AMD, as well as controls, conducted at 20 clinical sites across Europe (Denmark, France, Germany, Italy, Netherlands, Portugal, and United Kingdom).[20,21] Patients self-completed the VILL and EQ-5D-5L questionnaires unless they requested interviewer administration.[21] The measures were administered in different languages across the different countries (Danish [Denmark], Dutch [The Netherlands], English [United Kingdom], French [France], German [Germany], Italian [Italy], and Portuguese [Portugal]). The MACUSTAR study has been registered on clinicaltrials.gov under NCT03349801. We used the MACUSTAR baseline data (n = 301) collected April 2018 to February 2020, accessed 19th March 2021.

Analyses

The classification system was derived to retain the 3 factors from the VILL-33 (reading and accessing information, mobility and safety, and emotional well-being) to ensure it retains the dimensions deemed important to patients in its development[16] but also building upon the psychometric analyses of the measure already undertaken.[17] A parsimonious number of items within each factor was selected based on their overall performance using a range of psychometric analyses including Rasch analysis (an approach used previously, eg, Mukuria et al[22] and Rowen et al[23,24]), published criteria around suitability for inclusion in a PWM[25] alongside instrument developer views. Rasch analysis converts categorical item responses to points on a continuous latent scale using a logit model[26] and is informative for indicating better performing items within a factor.[27-29]

Psychometric performance was assessed using Stata version 15 by known-group validity, floor and ceiling effects, missing data, proportion of data on the response option "Didn't do this for other reasons/Does not apply to me," and correlation with the factor score, as well as Rasch analysis using Winsteps using the polytomous rating scale model.[17,30] Items that have known-group validity and hence reflect differences across severity groups are preferred, where this is captured using impairment (no and early AMD are merged and compared with the intermediate AMD group). Items with low ceiling and floor effects are preferred because the item cannot capture health improvement at the ceiling or health deterioration at the floor. Items with low levels of missing data are preferred, because utilities cannot be generated when there is missing data for any item. Items with lower proportions of responses for the response option "Didn't do this for other reasons/Does not apply to me" are preferred because they must be treated as missing data when utilities are generated. Items are preferred with higher correlations with the factor score because the classification system will contain the minimum number of items required in each factor to capture the range of severity and aspect of HRQoL reflected in the factor. Item correlations are also used to indicate where one of the highly correlated items can be selected (strong correlation $\geq$0.5, moderate correlation $<$0.5 to $\geq$0.3, and weak correlation $<$0.3).

Rasch models were separately estimated for each factor and used to inform the selection of a minimum number of items per factor using the following assessments:

- Item infit and outfit (based on unweighted mean squared residual values): values between 0.5 and 1.5 to indicate better fitting items within each factor, which is preferred.

- Spread: using ability across item response categories 1 to 4, which indicates discrimination, for each item within the factor, with larger spread preferred.
- Ordered thresholds using category probability curves: this indicates whether the response options are ordered correctly to capture increasing severity, in which items that have ordered response options are preferred.
- Differential item functioning (DIF): by age and sex, to indicate whether the items perform similarly for patients with different sex (male vs female) or different age group (≤70 vs >70), contrast ≥0.64 logits,[31] in which significance is assessed using the Mantel-Haenszel *P* value. The investigation was focused on uniform DIF. Items with no DIF are preferred.

Valuation of the VILL-UI Classification System

Valuation method and population

Discrete choice experiment (DCE) was used, in which a participant is asked which of 2 health-state descriptions they prefer. Each health-state description is generated using the VILL-UI classification system, with a description of the severity level of each item in the classification system, plus an attribute that reflects how long the health state lasts in years. This technique is often called DCE with duration, or DCE-TTO, because it combines duration with the health state and is increasingly used in health-state valuation studies.[32-34] Duration levels of 1, 4, 7, and 10 years were chosen to reflect a range of years that are plausible for the task and in accordance with previous studies (for example, Mulhern et al,[35] Norman et al,[36,37] and Rowen et al[23,38]). The inclusion of duration enables the results to be modeled to generate a health-state utility value for every health state described by the classification system, such that 0 is equivalent to dead, 1 is equivalent to full health, and a value below 0 means the health state is so bad that it is regarded as worse than being dead.[36,39]

Members of the public were selected to value the measure, as recommended by the National Institute of Health and Care Excellence (NICE) in England and Wales.[40] Sample size of 1000 per country was selected to ensure each choice set was valued at least 20 times with at least 1 choice set per parameter estimated in the regression.[41]

Valuation design

The DCE survey consists of a number of choice sets, ie, combinations of 2 health states with duration included. There are too many possible combinations to be able to select all possible combinations of health states and duration levels. In addition, 2 items in the classification system capturing mobility and safety were highly correlated and hence could not be treated as independent when selecting choice sets (ie, a health state would not make sense if 1 mobility and safety item had no difficulty and the other had a lot of difficulty); therefore, these items were merged into a single attribute. Choice sets were selected to ensure that the required model would be able to be estimated from survey responses using the d-create add-in command in Stata. This generates a D-efficient design and uses the modified Federov algorithm.[42] In total 80 choice sets were selected, which were 80 pairs of health states with a duration level.

The survey

The valuation survey was developed in both English and German to enable the survey to be administered online in the United Kingdom and Germany (where there is substantial interest in use of the measure). The VILL items used were obtained from the original instrument, in which the translation and cultural adaptation process followed established standards.[43] No prior knowledge of AMD was required to be able to understand or complete the survey, and participants were not informed that the health states were AMD-specific, although because of the nature of the health states, they were informed that the aspects of health and quality of life affected were because of their eyesight. The survey consisted of an information sheet and consent form; questions about the participant (ie, sociodemographic questions), their general health using EQ-5D-5L (UK scoring[44] and German scoring[45]) and their vision using the VILL-UI classification system; explanation of the DCE task, a practice DCE task (a dominant question with 1 clearly better answer), and 10 DCE questions (9 randomly selected from the 80 choice sets chosen statistically, ie, 18 health states, and 1 dominant question with 1 clearly better answer to assess participant understanding, see Fig. 1); and questions about the understanding/difficulty of the survey. For the German survey, additional questions were included around medical history related to eye problems and health services, with these results to be reported elsewhere. The survey was soft launched with 100 participants in each country and the data checked before proceeding to recruit the remaining participants. The survey was hosted by surveyengine who also managed participant recruitment from an existing online panel, with participants recruited according to combined quotas for age and sex to ensure a representative sample.

The UK survey received approval from University of Sheffield Research Ethics Committee administered by the School of Health and Related Research (approval ID 047001) and the German

Figure 1. Example discrete choice experiment survey screenshot of the dominance question.

	Life A	Life B
	You live for **10 years** with the following then you die:	You live for **10 years** with the following then you die:
	No difficulty recognizing small objects in dim lighting	**A lot** of difficulty recognizing small objects in dim lighting
	A little difficulty reading print against a colourful background	**A lot** of difficulty reading print against a colourful background
	No difficulty seeing steps or curbs in the darkt	**A lot** of difficulty seeing steps or curbs in the dark
	Never feel unsafe as a pedestrian or cyclist at dawn or at night	**Often** feel unsafe as a pedestrian or cyclist at dawn or at night
	Sometimes feel worried that your eyesight might get worse	**Often** feel worried that your eyesight might get worse
Which do you prefer?	○	○

Table 1. Summary of psychometric and Rasch analysis performance of VILL-33 items for selection in classification system.

Item no.	Item wording	Item performs well across Peasgood criteria	Known-group validity	Ceiling effect (% response at best response option)	Item data for response option 5, did not do for other reasons	Correlation	Differential item functioning (sex, age)	Spread	Ordered thresholds	Item infit and outfit	Item performs well psycho metrically
Reading and accessing information											
1	Adjusting to the dark when entering a dimly lit room? (eg, a restaurant at night)		0.005	55.5	2.3	0.641	0.382, 0.610	−1.28, 2.98	Yes	1.13, 1.04	✓
2	Recognizing small objects in dim lighting? (eg, coins)	✓	0.007	39.2	0.7	0.718	0.108, 0.852	−2.34, 3.31	Yes	1.12, 1.13	✓
3	Recognizing people's faces outside during dusk?	✓	0.003	50.5	0.3	0.701	0.632, 0.834	−3.14, 3.12	Yes	1.07, 1.10	✓
4	Recognizing people or objects by candlelight?		0.002	47.3	9.7	0.671	0.008, 0.838	−2.89, 3.16	Yes	1.03, 1.12	
5	Seeing things clearly close up in the middle of your field of vision?		0.010	69.1	0.3	0.546	<0.001, 0.821	−3.78, 2.75	Yes	1.32, 1.33	
6	Reading print which has a low contrast to its background?		<0.001	30.6	0.3	0.788	0.440, 0.087	−2.65, 3.59	Yes	0.79, 0.80	✓
7	Reading print which is not black? (eg, gray)	✓	0.007	41.2	1.0	0.781	0.040, 0.026	−2.80, 3.43	Yes	0.81, 0.81	
8	Reading text on a digital display? (eg, in the car, on an electronic radio)	✓	0.180	65.1	1.7	0.655	0.400, 0.644	−3.39, 2.80	Yes	1.23, 1.35	
9	Reading print against a colorful background? (eg, a brochure)	✓	0.008	55.2	0.3	0.662	0.419, 0.108	−3.40, 3.09	Yes	1.07, 1.01	✓
10	Reading a paperback novel in dim lighting?		<0.001	23.3	6.0	0.834	0.901, 0.669	−2.06, 4.03	Yes	0.65, 0.63	✓
11	Reading a newspaper in dim lighting?		<0.001	24.9	3.3	0.829	0.760, 0.417	−2.13, 4.04	Yes	0.66, 0.64	✓
12	Reading a menu in a dimly lit restaurant?		<0.001	36.5	2.0	0.830	0.223, 0.291	−2.63, 3.66	Yes	0.79, 0.74	✓
13	Reading labels or instructions on medicine bottles in good lighting?	✓	0.003	59.8	0.3	0.663	0.845, 0.549	−2.89, 2.98	Yes	1.32, 1.24	✓
14	Reading labels or instructions on medicine bottles in dim lighting?	✓	<0.001	16.0	2.7	0.810	0.014, 0.218	−1.28, 4.07	Yes	0.81, 0.81	
15	Reading package labels or price tags in a shop?	✓	0.004	60.5	0.3	0.655	0.292, 0.723	−3.36, 3.07	Yes	1.04, 1.01	✓
28	Felt exhausted by reading in dim light?	✓	<0.001	22.0	6.7	0.742	0.027, 0.010	−0.13, 3.54	Yes	1.18, 1.18	
29	Needed additional lighting to see or read anything?	✓	<0.001	30.2	1.3	0.698	0.571, 0.628	−0.11, 3.70	Yes	1.26, 1.19	✓

continued on next page

Table 1. Continued

Item no.	Item wording	Item performs well across Peasgood criteria	Known-group validity	Ceiling effect (% response at best response option)	Item data for response option 5, did not do for other reasons	Correlation	Differential item functioning (sex, age)	Spread	Ordered thresholds	Item infit and outfit	Item performs well psycho metrically
Mobility and safety											
16	Driving a car on a sunny day? (with or without sunglasses)		0.918	47.2	19.6	0.565	0.588, 0.151	−2.50 to 3.19	Yes	1.22, 1.23	
17	Driving a car along a road lined with trees on a sunny day? (with or without sunglasses)		0.066	44.9	21.9	0.666	0.930, 0.188	−3.06 to 3.28	Yes	1.03, 0.96	
18	Driving a car at night?		0.001	26.9	22.6	0.830	0.082, 0.418	−1.76 to 4.15	Yes	0.84, 0.80	✓
19	Driving a car at night in the rain?		0.002	18.6	22.9	0.866	0.579, 0.904	−1.36 to 4.60	Yes	0.80, 0.80	✓
20	Reading street signs in time when driving by?		0.102	51.8	19.3	0.630	0.055, 0.689	−2.99 to 3.15	Yes	1.29, 1.26	
21	Walking on uneven ground in the dark?	✓	0.013	34.9	4.3	0.693	0.627, 0.030	−3.74 to 3.59	Yes	0.92, 0.97	
22	Going out to do things during dusk? (eg, visiting the supermarket or shops)	✓	0.052	70.4	1.7	0.526	0.947, 0.008	−4.37 to 2.86	Yes	1.04, 0.99	
23	Seeing steps or curbs in the dark?	✓	0.028	35.9	2.0	0.757	0.700, 0.152	−3.86 to 3.78	Yes	0.76, 0.81	✓
24	Getting your bearings in dimly lit or dark unfamiliar places?		0.069	32.6	3.0	0.749	0.354, 0.328	−3.43 to 3.89	Yes	0.80, 0.83	✓ (maybe)
25	Felt blinded by oncoming cars at night?	✓	0.059	15.0	14.3	0.728	0.847, 0.033	−0.29 to 4.25	Yes	1.11, 1.08	
26	Felt blinded by the sun while driving a car? (with or without sunglasses)		0.025	16.9	21.6	0.669	0.888, 0.725	−0.99 to 4.03	Yes	1.09, 1.09	✓
27	Felt unsafe as a pedestrian or cyclist at dawn or at night?		0.043	57.5	3.7	0.626	0.951, 0.859	−2.07 to 3.15	Yes	1.40, 1.13	✓
Emotional well-being											
30	Felt worried that your eyesight might get worse?	✓	0.022	28.9	0.7	0.849	0.106, 0.733	−1.75, 3.79	Yes	1.35, 1.38	✓
31	Felt worried about losing your independence?	✓	0.356	40.9	0.7	0.909	0.707, 0.118	−3.04, 3.96	Yes	0.87, 0.81	
32	Felt worried about the future?	✓	0.103	40.9	1.0	0.900	0.008, 0.393	−2.93, 4.00	Yes	0.86, 0.83	
33	Felt worried that your lifestyle might change due to your eye condition?	✓	0.042	34.9	5.0	0.905	0.330, 0.057	−3.21, 4.04	Yes	0.84, 0.82	

Note. Correlation has been generated between a dimension score (generated by summing the VILL-33 items for that dimension) and the item using Spearman rank correlation coefficient. Known-group validity has been assessed using the Wilcoxon-Mann-Whitney test in Stata for 2 groups: no and early AMD vs intermediate AMD, due to the sample sizes across the different severity groups of AMD, and the *P*-value is reported. DIF has been calculated using Table 30.1 in Winsteps, and the Mantel-Haenszel *P*-value reported, for DIF by sex (male, female) and age group (≤70, >70. Spread has been calculated using Table 13.3 in Winsteps that reports ability across item response categories, and the lowest number reported is for response option 1 and the highest number reported is for response option 4. Infit has been calculated using the unweighted mean-square statistics. The cut-off used in[16] was Items showing outfit or infit > mean-square value 1.4 were removed. The DIF for the emotional well-being dimension was generated including all 4 items in the model, but note that item 34 suffers from suffers from misfit to the model. AMD indicates age-related macular degeneration; DIF, differential item functioning; VILL-33, Vision Impairment in Low Luminance-33.

Table 2. Classification system for the DCE design.

VILL domain	VILL item	VILL-UI dimension	Level	Description
Reading and accessing information	2. Recognizing small objects in dim lighting (eg, coins)	Accessing information	1	No difficulty recognizing small objects in dim lighting
			2	A little difficulty recognizing small objects in dim lighting
			3	A lot of difficulty recognizing small objects in dim lighting
			4	Cannot recognize small objects in dim lighting
	9. Reading print against a colorful background (eg, a brochure)	Reading	1	No difficulty reading print against a colorful background
			2	A little difficulty reading print against a colorful background
			3	A lot of difficulty reading print against a colorful background
			4	Cannot read print against a colorful background
Mobility and safety	23. Seeing steps or curbs in the dark	Mobility and safety	1	No difficulty seeing steps or curbs in the dark (23,1) Never feel unsafe as a pedestrian or cyclist at dawn or at night (27,1)
	27. Feel unsafe as a pedestrian or cyclist at dawn or at night		2	A little difficulty seeing steps or curbs in the dark (23,2) Sometimes feel unsafe as a pedestrian or cyclist at dawn or at night (27,2)
			3	A little difficulty seeing steps or curbs in the dark (23,2) Often feel unsafe as a pedestrian or cyclist at dawn or at night (27,3)
			4	A little difficulty seeing steps or curbs in the dark (23,2) Always feel unsafe as a pedestrian or cyclist at dawn or at night (27,4)
			5	A lot of difficulty seeing steps or curbs in the dark (23,3) Often feel unsafe as a pedestrian or cyclist at dawn or at night (27,3)
			6	A lot of difficulty seeing steps or curbs in the dark (23,3) Always feel unsafe as a pedestrian or cyclist at dawn or at night (27,4)
			7	Cannot see steps or curbs in the dark (23,4) Often feel unsafe as a pedestrian or cyclist at dawn or at night (27,3)
			8	Cannot see steps or curbs in the dark (23,4) Always feel unsafe as a pedestrian or cyclist at dawn or at night (27,4)
Emotional well-being	30. Feel worried that your eyesight might get worse	Worry	1	Never feel worried that your eyesight might get worse
			2	Sometimes feel worried that your eyesight might get worse
			3	Often feel worried that your eyesight might get worse
			4	Always feel worried that your eyesight might get worse
		Duration	1	1 y
			4	4 y
			7	7 y
			10	10 y

Note. For items 23 and 27 level 1 = no difficulty; 2 = a little difficulty; 3 = a lot of difficulty; 4 = cannot do.
DCE indicates discrete choice experiment; VILL, Vision Impairment in Low Luminance; VILL-UI, Vision Impairment in Low Luminance - Utility Index.

survey received approval from the Human Research Ethics Committee of the University of Bonn (approval ID 255/22).

Analysis

The health and socioeconomic characteristics of the sample were assessed using descriptive statistics and compared with the country general public. Understanding and difficulty of the survey was assessed using self-reported responses and responses to the practice and dominance questions.

The DCE data were analyzed using Stata version 15 with choice as the dependent variable using the conditional logit model[36,39] with the specification:

$$\mu_{ij} = \alpha_i + \beta_1 t_{ij} + \beta_2' \mathbf{x}_{ij} t_{ij} + \varepsilon_{ij}$$

in which μ_{ij} is utility of individual i for health-state j, α_i is an individual specific constant term, t is life-years, ε_{ij} represents the error term, β_1 is the coefficient for duration, and β_2' represents the coefficients on the interaction terms of duration and severity levels of each item (level 1 is the reference level). Responses to the dominance question were not included in the modeling because it was not selected as part of the DCE design. The conditional logit was selected over mixed logit because the purpose was to generate average preference weights for each country's participants, which would be used in population-level outcomes evaluations. In this context, modeling unobserved heterogeneity was not a priority.

The marginal rate of substitution was used to generate the utility decrement (on the utility scale required to generate QALYs) for each

Table 3. Valuation samples of UK and German DCE survey respondents.

Sociodemographic characteristics *(German wording in brackets and italics if different)*		UK sample (n = 1004), %*	German sample (n = 1008), %*	UK general population, %†	German general population, %‡
Sex	Male	49.8	49.7	48.9	49.0
	Female	48.9	49.6	51.1	51.0
	Other/prefer not to say	1.3	0.7		
Age	Mean age (SD)	47.8 (17.5)	50.0 (16.4)		
	Age 18-44	42.0	38.3	43.7§	37.9‖
	45-64	32.8	36.8	32.7	35.9
	65+	21.2	24.9	23.7	26.2
	Prefer not to say	4.0	0.5		
Education	Education continued after age of 16	75.5	93.9		
	Degree or equivalent professional qualification	57.2	50.2		
Employment status	In employment or self-employment	59.2	56.3	61.7	60.7
	Retired	19.7	27.0	13.9	37.1
	Student	5.2	4.3	4.3	
	Long-term sick	3.9	2.1		
	Carer or volunteer *(Caring for family)*	1.4	1.0	4.4	
	Not seeking work	2.7	3.7		
	Unemployed	5.2	4.9	9.3	2.2
	Prefer not to say	1.5	0.4	4.3	
	Other	1.3	0.5	2.2	
Health					
General health	Excellent	13.0	5.3		
	Very good	30.0	22.8		
	Good	32.8	43.4		
	Fair	20.4	24.5		
	Poor	6.1	4.1		
EQ-5D-5L	Mean (SD)	0.75 (0.25)	0.84 (0.22)		
EQ-VAS	Mean (SD)	70.4 (21.3)	69.2 (22.4)		
VILL-UI					
Difficulty recognizing small objects in dim lighting (eg, coins)	Can't do because of eyesight	4.4	2.8		
	A lot	13.5	9.3		
	A little	33.0	35.0		
	None	47.3	50.7		
	Don't do this for other reasons	1.9	2.2		
Difficulty reading print against a colorful background (eg, a brochure)	Can't do because of eyesight	3.1	1.6		
	A lot	16.1	6.7		
	A little	29.1	24.3		
	None	49.8	65.0		
	Don't do this for other reasons	1.9	2.5		

continued on next page

Table 3. Continued

Sociodemographic characteristics *(German wording in brackets and italics if different)*		UK sample (n = 1004), %*	German sample (n = 1008), %*	UK general population, %†	German general population, %‡
Seeing steps or curbs in the dark	Can't do because of eyesight	2.9	1.9		
	A lot	11.8	7.3		
	A little	24.3	26.5		
	None	58.0	60.9		
	Don't do this for other reasons	3.1	3.4		
Feel unsafe as a pedestrian or cyclist at dawn or at night	Always	4.3	2.6		
	Often	6.9	8.3		
	Sometimes	17.2	25.6		
	Never	62.9	54.6		
	Don't do this for other reasons	8.8	8.9		
Feel worried that your eyesight might get worse	Always	6.3	3.0		
	Often	14.6	10.9		
	Sometimes	38.8	37.5		
	Never	39.6	45.4		
	Don't do this for other reasons	1.7	3.2		
Understanding and engagement					
Difficulty to answer DCE tasks *(Difficulty of DCE tasks)*	Very difficult to answer *(Very difficult)*	3.9	3.9		
	Quite difficult to answer *(Difficult)*	28.4	16.9		
	Neither difficult nor easy to answer *(Neither difficult nor easy)*	24.3	37.9		
	Fairly easy to answer *(Easy)*	30.0	30.6		
	Very easy to answer *(Very easy)*	13.5	10.8		
Practice question	Selected dominant option	88.4	84.3		
Dominance question	Selected dominant option	85.9	87.0		
Selected non-dominant option in both practice and dominance question		4.6	5.6		
Time in minutes taken to complete survey, mean (SD)		10.3 (12.5)	15.2 (12.2)		

DCE indicates discrete choice experiment; UK, United Kingdom; VILL-UI, Vision Impairment in Low Luminance - Utility Index.
*The UK survey was conducted in August 2022. The German survey was conducted in August and September 2022.
†Statistics for UK for age and gender are from the Office for National Statistics' Mid-Year Population Estimates June 2020. The statistics on employment status are for England in the Census 2011. The census includes persons aged 16 and above, whereas this study only surveys persons aged ≥18.
‡Statistics for Germany from Research Data Centre of the Federal Statistical Office and Statistical Offices of the Federal States, micro-census 2021.
§Age distribution is here reported as the percentage of all adults aged ≥18.
‖Age range 20 to 44 years because of availability of quota data.

Table 4. Regression analysis of UK and German DCE data and (anchored) preference weights.

Variable	UK	Germany		VILL-UI dimension	Level	UK (anchored) preference weights	German (anchored) preference weights (using fully consistent model)
	Standard model	Standard model	Fully consistent model				
Info2_LY	−0.038*	−0.021*	−0.023*	Accessing information	2	−0.076	−0.057
	(<0.001)	(0.001)	(0.001)				
Info3_LY	−0.095*	−0.087*	−0.078*		3	−0.189	−0.195
	(<0.001)	(<0.001)	(<0.001)				
Info4_LY	−0.127*	−0.069*	−0.078*		4	−0.253	−0.195
	(<0.001)	(<0.001)	(<0.001)				
Read2_LY	−0.019*	−0.024*	−0.023*	Reading	2	−0.038	−0.057
	(0.006)	(<0.001)	(0.001)				
Read3_LY	−0.081*	−0.074*	−0.07*		3	−0.161	−0.175
	(<0.001)	(<0.001)	(<0.001)				
Read4_LY	−0.131*	−0.085*	−0.082*		4	−0.261	−0.204
	(<0.001)	(<0.001)	(<0.001)				
Mob2_LY	−0.041*	−0.047*	−0.047*	Mobility and safety	2	−0.082	−0.117
	(<0.001)	(<0.001)	(<0.001)				
Mob3_LY	−0.093*	−0.098*	−0.099*		3	−0.185	−0.247
	(<0.001)	(<0.001)	(<0.001)				
Mob4_LY	−0.124*	−0.117*	−0.115*		4	−0.247	−0.287
	(<0.001)	(<0.001)	(<0.001)				
Mob5_LY	−0.137*	−0.159*	−0.158*		5	−0.273	−0.394
	(<0.001)	(<0.001)	(<0.001)				
Mob6_LY	−0.170*	−0.188*	−0.187*		6	−0.339	−0.466
	(<0.001)	(<0.001)	(<0.001)				
Mob7_LY	−0.164*	−0.168*	−0.169*		7	−0.327	−0.421
	(<0.001)	(<0.001)	(<0.001)				
Mob8_LY	−0.226*	−0.221*	−0.222*		8	−0.450	−0.554
	(<0.001)	(<0.001)	(<0.001)				
Worry2_LY	−0.011	−0.014†	−0.012†	Worry	2	−0.022	−0.030
	(0.121)	(0.039)	(0.048)				
Worry3_LY	−0.034*	−0.053*	−0.051*		3	−0.068	−0.127
	(<0.001)	(<0.001)	(<0.001)				
Worry4_LY	−0.060*	−0.094*	−0.092*		4	−0.120	−0.229
	(<0.001)	(<0.001)	(<0.001)				
LY	0.502*	0.402*	0.401*				
	(<0.001)	(<0.001)	(<0.001)				
Respondents	1004	1008	1008				
Observations	18 072	18 144	18 144				
Log likelihood	−5354	−5557	−5559				
Rho-squared	0.145	0.116	0.116				

Note. *P*-values are in parentheses.
DCE indicates discrete choice experiment; LY, life-year; UK, United Kingdom; VILL-UI, Vision Impairment in Low Luminance - Utility Index.
**P* values are significant at 1% level.
†*P* values are significant at 5% level.

severity level of each item[36,39] by dividing the coefficient of each severity level of each item by the coefficient of life-years, with standard errors calculated using the Delta method. The utility of each health state is generated by summing 1 plus the utility decrements.

Model performance was examined using sign, significance, and logical consistency of coefficients, (in which as health worsens, utility does not increase), log likelihood, and Rho-squared. Any inconsistent adjacent levels (in which health worsens but utility increases) were merged to produce a single utility decrement for the levels to enable the estimation of a fully consistent model, an approach used previously (eg, Mukuria et al,[22] Rowen et al, [23,24,38] and Norman et al[36,37]). Selection criteria for the preference weights model were logical consistency of coefficients and acceptable model performance (using sign and significance of coefficients, log likelihood, and Rho-squared).

Duration was modeled as a linear and continuous variable, and this assumption was examined through modeling duration as a categorical variable and plotting the coefficients to assess linearity.[46] Robustness of the results was examined by estimating models on subsets of respondents, to exclude participants in situations which their understanding or engagement may be questioned using their responses to the dominant question and practice (dominant) question and self-reporting understanding and difficulty of the DCE tasks.

Additional models were estimated to explore preference heterogeneity, in which preferences vary across respondents, which included interaction effects for various sociodemographic and health characteristics and the impact assessed on the sign and significance of the coefficients and size of utility decrement (using marginal rate of substitution as detailed above).

Results

VILL-UI Health-State Classification System Item Selection

Table 1 summarizes the results of the psychometric and Rasch analyses for each VILL-33 item, alongside whether they meet published criteria for item selection for a PWM[25] to the VILL-33. Details of the MACUSTAR sample are included in the supplemental materials (Appendix Table A1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.02.001), as well as a summary of the application of each of the published criteria (Appendix Table A2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.02.001), and correlations of the VILL-33 items (Appendix Table A3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.02.001). Table 2 details the selected classification system, which describes 512 health states.

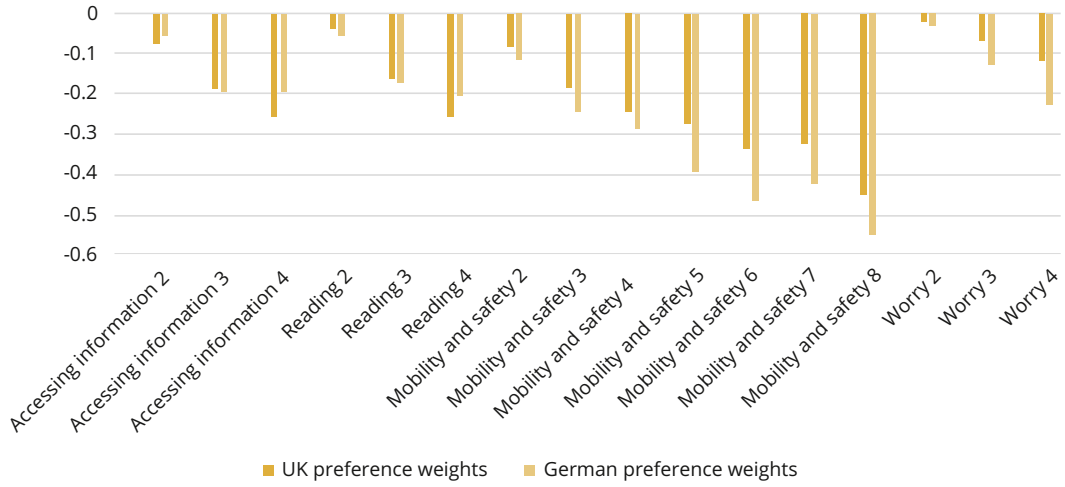
Reading and accessing information

Six items performed well on the published criteria and psychometrically, items 2, 3, 9, 13, 15, and 29 (see Table 1). Items 2 and 3 captured similar aspects with strong correlation, and item 2 performed marginally better. Items 9, 13, and 15 had large ceiling effects, and all had similar spread. Item 29 had spread concentrated at the milder impairment. Items 2 and 9 were selected to capture accessing information under low luminescence and reading, which are separate factors captured within this VILL factor. Item 2 performed well on the published criteria and psychometrically and is applicable to most people, with higher correlation with the domain score and lower ceiling effects than comparable items. Item 9 performed well on the published criteria and psychometrically, is a very relevant activity for day-to-day life that applies to everyone, and is arguably more relevant than similar items within the domain.

Mobility and safety

Only 1 item performed consistently well on the published criteria and psychometrically, item 23 (see Table 1). Seven of the items captured impact on driving, which are not applicable to all participants because not everyone drives and were not selected on this basis. Because the factor captures both mobility and safety (in relation to safe mobility) because these were qualitatively important aspects of this factor in the VILL development, item 23 was selected to represent mobility and item 27 to represent safety because this item 27 performed well psychometrically. Cross-tabulations of responses to these items indicated that the 2 items concur together and therefore should be merged into a single attribute for valuation (for an example, where this has been done previously, see Rowen et al[47]).

Figure 2. Plot of (anchored) preference weights per dimension and level in the UK and German value sets.



UK indicates United Kingdom.

Emotional well-being

All items performed well across the published criteria, although only 1 of the 4 items, item 30, performed well psychometrically and was selected to represent emotional well-being (see Table 1).

VILL-UI Valuation

No issues were identified after the survey soft launch, and data collection proceeded as planned with no changes made to the survey.

Valuation sample

The valuation samples (United Kingdom, n = 1004; Germany n = 1008) (see Table 3 and Appendix Table A4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.02.001) are representative of the country populations for age and sex, although there are some differences for employment status with larger proportion of retired individuals in the UK sample and a lower proportion in the German sample. All respondents completing the survey were included in the sample. Sample participants have a range of health severity, and a substantial proportion of participants experience some impairment to their vision (Table 3). Given the recruitment method, it was not possible to generate a response rate of people invited to participate.

Understanding and engagement

In the German and UK samples, 19.8% and 32.3% of participants, respectively, reported the DCE tasks were quite or very difficult to answer (see Table 3). Understanding and engagement was indicated by the responses to the practice question and dominant question (each had a dominant profile which was clearly better) (see Fig. 1 where the dominant task is shown as an example), in which between 84.3% and 88.4% (German and UK samples, respectively) of participants correctly chose the better option (this proportion is comparable to the DCE with duration valuation of another PWM of 86.6% to 90.1%[23]).

Regression analysis

Table 4 reports the modeled coefficients and the utility decrements generated using the marginal rate of substitution. The DCE analyses produced logically consistent and significant coefficients, meaning that as health deteriorated, the utility index accurately reflected this, and the model performed as expected. This was with the exception of 1 inconsistency in the modeled German data (for reading and accessing information levels 3 and 4), and a consistent model was therefore estimated for the German data (merging reading and accessing information levels 3 and 4 to a single dummy variable). The standard model for the United Kingdom and the consistent model for Germany should be used as VILL-UI preference weights. There is no inconsistency in the coefficients observed between severity levels 6 and 7 in the mobility and safety dimension because these are made up of 2 items and levels 6 and 7 have 1 item at the most severe level and the other item at the third most severe level.

Plots of duration when modeled as a continuous variable indicate that it is appropriate to assume linearity for the duration variable. Robustness analyses indicated that the exclusion of participants who may not have engaged or understood had minimal impact on the modeled results (Appendix Tables A5 and A6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.02.001). Assessments of preference heterogeneity found that men typically had lower preference weightings, ie, smaller utility decrement) for the mobility and safety factor but otherwise there was no clear pattern indicating preferences across different groups of participants according to their observable characteristics (Appendix Tables A7 and A8 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.02.001).

Each utility decrement is summed to 1 to generate the overall utility value for the health state. For example, for a health state with accessing information at level 4, reading at level 3, mobility and safety at level 8, and worry at level 1, the UK utility value is = 1-0.253-0.161-0.450-0 = 0.136; the German utility value is 1-0.195-0.175-0.554-0 = 0.076.

Comparison of UK and German VILL-UI preference weights

Figure 2 plots the UK and German VILL-UI preference weights. For the reading and accessing information items, the preference weights are similar for United Kingdom and Germany, although are larger for the UK weights for the most severe level (level 4). The German preference weights are larger for the mobility and safety factor and the emotional well-being factor (the worry item) in comparison with the UK weights.

Discussion

This study has generated an AMD-specific PWM with UK and German preference weights. The results of the valuation survey enable utility values to be generated for every health state defined by the VILL-UI classification system, with different preference weights generated from UK and German samples, to enable the VILL to be used to inform economic evaluation using cost-utility analysis.

Although the UK and German preference weights are similar, there is a higher relative weighting to mobility, safety and emotional well-being in the German weights in comparison with the UK weights, and the utility for the worst state is lower for the German weights (−0.182 in comparison with −0.084). Differences in preferences across countries are expected because sample compositions vary across countries, and relative preferences across dimensions and severity levels differ according to the socioeconomic characteristics and cultures across countries. Further models examining heterogeneity, for example, latent class analyses, have not been undertaken because cost-utility analyses are undertaken at the average population level, meaning that the utility decrements reported here are those that are appropriate for economic evaluation.

The study elicited preferences from the general population in accordance with recommendations from NICE.[40] However, for Germany cost-utility analysis is not explicitly recommended for use,[48,49] meaning that there is no clear German guidance on health-state utilities. Some countries, for example, Sweden, recommend using the preferences of patients.[50]. It would be anticipated that the preferences of patients could deviate from the preferences of the general public.

The AMD-specific VILL-UI can be used for a variety of purposes, including examinations of how health changes over time, comparisons of health across populations, healthcare providers, and treatments, as well as to generate QALYs. For health technology assessment, many agencies require the use of a generic PWM (for example, Rowen et al[51]) and EQ-5D in particular, for example, NICE.[40] However, under recent NICE guidelines, a condition-specific PWM can be used when EQ-5D is inappropriate for a condition, and arguably AMD is 1 such condition.

One alternative to using a condition-specific PWM is to include EQ-5D with a vision bolt-on, an additional dimension reflecting visual impairment, and applying an appropriate preference

weighting.[52-55] However, there are no preference weights available for the EQ-5D-5L visual impairment bolt-on, meaning utilities for QALYs cannot be directly generated. Inclusion of a bolt-on dimension means that any increased sensitivity is due to a single item reflecting vision, which is unlikely to capture all that is important for people with visual impairment, and other dimensions are retained, which are not expected to be sensitive to changes in vision-related quality of life. There are considerable advantages of using an AMD-specific measure over mapping to EQ-5D (or another generic PWM) or using a vision bolt-on to EQ-5D, in that the VILL-UI contains the aspects of HRQoL deemed important for AMD patients from qualitative work and is not limited to a single item on vision. Consequently, a disadvantage of VILL-UI is that it does not capture wider HRQoL aspects of the patient because the utilities that are generated are vision specific. Examination of the psychometric performance of VILL-UI in comparison with EQ-5D-5L is recommended and will be conducted as part of ongoing MACUSTAR research in AMD.

Key advantages of DCE with duration are that participants consider trading years of life for improvements in aspects of HRQoL using a task considered easier to understand than time-trade-off, meaning it can be completed without an interviewer present. However, the task involves the consideration of a large amount of information simultaneously and therefore can be cognitively challenging to answer. The DCE with duration approach infers the position of dead but does not directly observe this in the tasks presented, which remains a criticism.[56] Other DCE variants have been explored in the literature that allow anchoring directly, such as a triplet with death, but their use is restricted to a small number of studies.[56]

Potential limitations of the approach used here include the assumption of linear time preference in the DCE survey, which could be questioned, although a plot of coefficients for duration entered as dummy variables for the levels indicated the relationship was linear. Furthermore, the DCE survey was conducted online with an existing panel of participants signed up to complete market research who may not be fully representative of the wider population and are not intended to be representative of patient preferences. As with all stated preference studies, the elicited preferences may not reflect revealed preferences. In the analyses used to select items for the classification system, there are additional analyses that could have been undertaken, but selection was based on a large range of criteria.

Study strengths include the development of the classification system via an international collaboration in AMD research as part of the MACUSTAR study involving instrument developers, clinicians, and health economists, who considered throughout the process the qualitative work with patients used to develop the parent VILL-33 measure.[16]

This study has generated the AMD-specific VILL-UI PWM for use in cost-effectiveness analyses and similar assessments. The VILL-UI can generate utilities from both existing and prospective VILL data using the preferences of the UK and German public.

Author Disclosures

Links to the disclosure forms provided by the authors are available here.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.02.001.

Article and Author Information

Accepted for Publication: February 8, 2024

Published Online: xxxx

doi: https://doi.org/10.1016/j.jval.2024.02.001

Author Affiliations: Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, England, UK (Rowen, Carlton, Wickramasekera, Brazier); Department of Ophthalmology, University of Bonn, Germany (Terheyden, Finger).

Correspondence: Donna Rowen, PhD, Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, England, United Kingdom. Email: d.rowen@sheffield.ac.uk

Author Contributions: *Concept and design:* Rowen, Carlton, Terheyden, Finger, Wickramasekera, Brazier
Acquisition of data: Rowen, Carlton, Terheyden, Finger, Wickramasekera, Brazier
Analysis and interpretation of data: Rowen, Carlton, Terheyden, Finger, Wickramasekera, Brazier
Drafting of the manuscript: Rowen
Critical revision of paper for important intellectual content: Rowen, Carlton, Terheyden, Finger, Wickramasekera, Brazier
Statistical analysis: Rowen
Obtaining funding: Rowen

Funding/Support: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116076. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. The communication reflects the author's view and neither IMI nor the European Union, EFPIA, or any Associated Partners are responsible for any use that may be made of the information contained therein.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors would like to thank all participants in the online survey and all patients who participated in the MACUSTAR study. MACUSTAR consortium members: H. Agostini, L. Altay, R. Atia, F. Bandello, P.G. Basile, C. Behning, M. Belmouhand, M. Berger, A. Binns, C.J.F. Boon, M. Böttger, C. Bouchet, J.E. Brazier, T. Butt, C. Carapezzi, J. Carlton, A. Carneiro, A. Charil, R. Coimbra, M. Cozzi, D.P. Crabb, J. Cunha-Vaz, C. Dahlke, L. de Sisternes, H. Dunbar, R.P. Finger, E. Fletcher, H. Floyd, C. Francisco, M. Gutfleisch, R. Hogg, F.G. Holz, C.B. Hoyng, A. Kilani, J. Krätzschmar, L. Kühlewein, M. Larsen, S. Leal, Y.T.E. Lechanteur, U.F.O. Luhmann, A. Lüning, I. Marques, C. Martinho, G. Montesano, Z. Mulyukov, M. Paques, B. Parodi, M. Parravano, S. Penas, T. Peters, T. Peto, M. Pfau, S. Poor, S. Priglinger, D. Rowen, G.S. Rubin, J. Sahel, D. Sanches Fernandes, C. Sánchez, O. Sander, M. Saßmannshausen, M. Schmid, S. Schmitz-Valckenberg, H. Schrinner-Fenske, J. Siedlecki, R. Silva, A. Skelly, E. Souied, G. Staurenghi, L. Stöhr, D. Tavares, J. Tavares, D.J. Taylor, J.H. Terheyden, S. Thiele, A. Tufail, M. Varano, L. Vieweg, J. Werner, L. Wintergerst, A. Wolf, N. Zakaria.

REFERENCES

1. Brooks R, Group EQ. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
2. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. 2nd ed. Oxford, United Kingdom: Oxford; 2016.
3. Fink DJ, Terheyden JH, Berger M, et al. The importance of visual health-a representative population survey. *Dtsch Ärztebl Int*. 2022;119(29-30):506–507.
4. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ*. 2018;19(4):557–570.
5. Tosh J, Brazier J, Evans P, Longworth L. A review of generic preference-based measures of health-related quality of life in visual disorders. *Value Health*. 2012;15(1):118–127.
6. Pennington BM, Hernández-Alava M, Hykin P, et al. Mapping from visual acuity to EQ-5D, EQ-5D with vision bolt-on, and VFQ-UI in patients with macular edema in the LEAVO trial. *Value Health*. 2020;23(7):928–935.

7. Au Eong KG, Chan EW, Luo N, et al. Validity of EuroQOL-5D, time trade-off, and standard gamble for age-related macular degeneration in the Singapore population. *Eye (Lond)*. 2012;26(3):379–388.
8. Choi S, Park SM, Jee D. Utility values for age-related macular degeneration patients in Korea. *PLoS One*. 2018;13(7):e0201399.
9. Cruess A, Zlateva G, Xu X, Rochon S. Burden of illness of neovascular age-related macular degeneration in Canada. *Can J Ophthalmol*. 2007;42(6):836–843.
10. Espallargues M, Czoski-Murray CJ, Bansback NJ, et al. The impact of age-related macular degeneration on health status utility values. *Invest Ophthalmol Vis Sci*. 2005;46(11):4016–4023.
11. Lotery A, Xu X, Zlatava G, Loftus J. Burden of illness, visual impairment and health resource utilisation of patients with neovascular age-related macular degeneration: results from the UK cohort of a five-country cross-sectional study. *Br J Ophthalmol*. 2007;91(10):1303–1307.
12. Payakachat N, Summers KH, Pleil AM, et al. Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Qual Life Res*. 2009;18(7):801–813.
13. Ruiz-Moreno JM, Coco RM, Garcia-Arumi J, Xu X, Zlateva G. Burden of illness of bilateral neovascular age-related macular degeneration in Spain. *Curr Med Res Opin*. 2008;24(7):2103–2111.
14. Soubrane G, Cruess A, Lotery A, et al. Burden and health care resource utilization in neovascular age-related macular degeneration: findings of a multicountry study. *Arch Ophthalmol*. 2007;125(9):1249–1254.
15. Kim J, Kwak HW, Lee WK, Kim HK. Impact of photodynamic therapy on quality of life of patients with age-related macular degeneration in Korea. *Jpn J Ophthalmol*. 2010;54(4):325–330.
16. Pondorfer SG, Terheyden JH, Overhoff H, Stasch-Bouws J, Holz FG, Finger RP. Development of the vision impairment in low luminance questionnaire. *Transl Vis Sci Technol*. 2021;10(1):S5.
17. Terheyden JH, Pondorfer SG, Behning C, et al. Disease-specific assessment of Vision Impairment in Low Luminance (VILL) in age-related macular degeneration – a MACUSTAR study report. *Br J Ophthalmol*. 2023;107(8):1144–1150.
18. Terheyden JH, Mekschrat L, Ost RAD, et al. Interviewer administration corresponds to self-administration of the Vision Impairment in Low Luminance (VILL) questionnaire. *Transl Vis Sci Technol*. 2022;11(4):21.
19. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–225.
20. Finger RP, Schmitz-Valckenberg S, Schmid M, et al. MACUSTAR: development and clinical validation of functional, structural, and patient-reported endpoints in intermediate age-related macular degeneration. *Ophthalmologica*. 2019;241(2):61–72.
21. Terheyden JH, Holz FG, Schmitz-Valckenberg S, et al. Clinical study protocol for a low-interventional study in intermediate age-related macular degeneration developing novel clinical endpoints for interventional clinical trials with a regulatory and patient access intention—MACUSTAR. *Trials*. 2020;21(1):1–11.
22. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a preference-based measure for myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846–855.
23. Rowen D, Powell P, Mukuria C, Carlton J, Norman R, Brazier J. Deriving a preference-based measure for people with Duchenne muscular dystrophy from the DMD-QoL. *Value Health*. 2021;24(10):1499–1510.
24. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health*. 2011;14(5):721–731.
25. Peasgood T, Mukuria C, Carlton J, Connell J, Brazier J. Criteria for item selection for a preference-based measure for use in economic evaluation. *Qual Life Res*. 2020;30(5):1425–1432.
26. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: University of Chicago Press; 1960.
27. Mavranezouli I, Brazier J, Young A, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of the CORE-6D from a measure of common mental health problems (CORE-OM). *Qual Life Res*. 2011;20(3):321–333.
28. Young T, Yang Y, Brazier J, Tsuchiya A. The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Med Decis Mak*. 2011;31(1):195–210.
29. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Qual Life Res*. 2009;18(2):253–265.
30. Boone WJ, Noltemeyer A, Yates G. Rasch analysis: a primer for school psychology researchers and practitioners. *Cogent Educ*. 2017;4(1):1416898.
31. Zwick R, Thayer DT, Lewis C. An empirical Bayes approach to Mantel-Haenszel DIF analysis. *J Educ Meas*. 1999;36(1):1–28.
32. Bahrampour M, Byrnes J, Norman R, Scuffham PA, Downes M. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. *Eur J Health Econ*. 2020;21(7):983–992.
33. Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *Pharmacoeconomics*. 2019;37(1):29–43.
34. Wang H, Rowen D, Brazier J, Jiang L. Discrete choice experiments in health state valuation: a systematic review of progress and new trends. *Appl Health Econ Health Policy*. 2023;21(2):405–418.
35. Mulhern BJ, Bansback N, Norman R, Brazier J, SF-6Dv2 International Project Group. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. *Med Care*. 2020;58(6):566–573.
36. Norman R, Viney R, Brazier J, et al. Valuing SF-6D health states using a discrete choice experiment. *Med Decis Mak*. 2014;34(6):773–786.
37. Norman R, Mercieca-Bebber R, Rowen D, et al. UK utility weights for the EORTC QLU-C10D. *Health Econ*. 2019;28(12):1385–1401.
38. Rowen DL, Mulhern B, Stevens K, Vermaire E. Estimating a Dutch value set for the paediatric preference-based CHU-9D using a discrete choice experiment with duration. *Value Health*. 2018;21(10):1234–1242.
39. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ*. 2012;31(1):306–318.
40. NICE health technology evaluations: the manual. National Institute for Health and Care Excellence. http://www.nice.org.uk/process/pmg36. Accessed July 3, 2023.
41. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics*. 2008;26(8):661–677.
42. DCREATE. *Stata* Module to Create Efficient Designs for Discrete Choice Experiments *[computer program]*. Boston College Department of Economics; 2015. http://econpapers.repec.org/RePEc:boc:bocode:s458059. Accessed July 3, 2023.
43. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94–104.
44. Hernández-Alava M, Pudney S, Wailoo AJ. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study. National Institute for Health Research (NIHR) policy Research Unit in economic methods of evaluation in health & social care interventions. Universities of Sheffield and York. Report no. 063. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/estimating-the-relationship-betweenE-Q-5D-5L-and-EQ-5D-3L.pdf; 2020. Accessed July 3, 2023.
45. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663–674.
46. Payne K, Fargher EA, Roberts SA, et al. Valuing pharmacogenetic testing services: a comparison of patients' and health care professionals' preferences. *Value Health*. 2011;14(1):121–134.
47. Rowen D, Wickramasekera N, Hole A, Keetharuth D, Wailoo A. A DCE to elicit general population preferences around the factors influencing the choice to make clinical negligence claims. *Value Health*. 2022;25(8):1404–1415.
48. Fricke F, Dauben HP. Health technology assessment: a perspective from Germany. *Value Health*. 2009;12(2 suppl 2):S20–S27.
49. Institute for Quality and Efficiency in Health Care (IQWiG). Health Technology Assessment: A Perspective From Germany. Köln, Germany: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen.
50. Pharmaceutical Benefits Board. *General Guidelines for Economic Evaluations From the Pharmaceutical Benefits Board*. Sweden: Pharmaceutical Benefits Board; 2003.
51. Rowen DL, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics*. 2017;35(suppl 1):11–19.
52. Gandhi M, Ang M, Teo K, et al. A vision 'bolt-on' increases the responsiveness of EQ-5D: preliminary evidence from a study of cataract surgery. *Eur J Health Econ HEPAC Health Econ Prev Care*. 2020;21(4):501–511.
53. Luo N, Wang X, Ang M, et al. A vision "bolt-on" item could increase the discriminatory power of the EQ-5D index score. *Value Health*. 2015;18(8):1037–1042.
54. Haywood P, Sampson C, Addo R, et al. Development Of EQ-5D-5L bolt-ons for cognition and vision. *Value Health*. 2019;22:S733. S733.
55. Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value Health*. 2015;18(1):52–60.
56. Norman R, Mulhern B, Lancsar E, et al. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. *Pharmacoeconomics*. 2023;41(4):427–438.